

Using Sound to Enhance Users' Experiences of Mobile Applications

Mats Liljedahl
Interactive Institute
Acusticum 4
SE-941 28 Piteå, Sweden
+46 70 540 58 88
mats.liljedahl@tii.se

Nigel Papworth
Interactive Institute
1st line of address
2nd line of address
+46 70 662 79 00
nigel.papworth@tii.se

ABSTRACT

The latest smartphones with GPS, electronic compass, directional audio, touch screens etc. hold potentials for location based services that are easier to use compared to traditional tools. Rather than interpreting maps, users may focus on their activities and the environment around them. Interfaces may be designed that let users search for information by simply pointing in a direction. Database queries can be created from GPS location and compass direction data. Users can get guidance to locations through pointing gestures, spatial sound and simple graphics. This article describes two studies testing prototypic applications with multimodal user interfaces built on spatial audio, graphics and text. Tests show that users appreciated the applications for their ease of use, for being fun and effective to use and for allowing users to interact directly with the environment rather than with abstractions of the same. The multimodal user interfaces contributed significantly to the overall user experience.

Categories and Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI):
Miscellaneous.

General Terms

Performance, Design, Reliability, Experimentation, Human Factors, Verification.

Keywords

Multimodal user interface, spatial audio, GPS, electronic compass, tourist guide, pervasive game

1. INTRODUCTION

When visiting new places you frequently want information about restaurants, shopping, places of historic interest etc. As a tourist you want to experience the place you are visiting. You want to see, hear, smell and generally use your senses to take in the environment. Smartphones are potentially perfect tools for delivering such location-based information and lately many new applications have also been released to take advantage of some of these new capabilities.

Visual maps must be interpreted in order to be of use. Interpreting

a map and relating it to the physical surroundings is a relatively demanding task [1]. Moreover, maps often require the users' full visual attention, disrupt other activities and may weaken the users' perception of the surroundings. All in all, maps are in many ways demanding tools for navigation.

In recent years totally new types of graphical representations of geographic information have emerged. One example is 360 degrees panorama views on street level of whole cities [2]. Other applications utilize Augmented Reality (AR) to correlate abstract representations to the surroundings. These types of tools are less abstract than traditional maps, are more concrete and demand less interpretation in order to be useful. But they still, to a large degree, require the users' full visual attention.

A problem with using the new technologies to support travelling experiences is that users run the risk to get absorbed by texts, maps, videos etc. on the device screen, and thus get alienated from the surroundings and the travelling experience they seek. This in turn may lead to that the users get second-hand experiences of the environment they are visiting, rather than the sought-after first-hand experience.

In this paper we will concentrate on two facets of this problem. The first is that the device screen contains so much of the total amount of travel information that the users might find themselves looking more at the screen than on the surroundings. The second facet is that interpreting more or less abstract information on maps, texts, images etc. may take up significant shares of the users' overall cognitive resources, leaving little room to actually experience the new environment.

The work presented here explored ways to overcome the problem by designing multimodal user interfaces based on users' everyday abilities such as directional hearing and point and sweep gestures. Today's smartphones know where you are, in what direction you are pointing the device and they have systems for rendering spatial audio. These readily available technologies hold the potential to make information more easy to interpret and use, demand less cognitive resources and free the users from having to look more or less constantly on a device screen. Potentially the problems could be overcome by designing user interfaces that combine sound, graphics, pointing gestures and geographic location into multimodal user interfaces working in very much the same way as the user herself does naturally in her everyday life.

In the project, two prototypic applications were designed, one tourist guide and one gaming application. Both applications had multimodal user interfaces based on GPS positioning and electronic compass as main input from the users, and a balanced mix of spatial and non-spatial audio, text and graphics as output to the users. The guide app allowed users to search for points of

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AM '12, September 26 - 28 2012, Corfu, Greece

Copyright 2012 ACM 978-1-4503-1569-2/12/09 \$15.00.

interest (POIs) and to get guidance to selected locations. In the game application, the users' movements in the physical world were reflected in the game's virtual world. Information from the virtual game world about directions and distances to opponents was conveyed through audio.

2. BACKGROUND

This study was inspired and guided by four concepts and aims: minimal attention user interfaces, eyes-free interaction, decreased cognitive loads on the users, and aesthetics of interaction. The study was also inspired by and based on a number of previous research efforts from several disciplines, including computer games, electronic navigation and ubiquitous computing.

Modes like auditive or haptic senses have been used for navigation applications in several studies. Examples include Tsukada and Yasymua [3], Frey [4], Amemiya et al. [5], Spath [6], Loomis [7], Kramer et al. [8], and Evett et al. [9]. But, as is often the case, the visual modality has drawn most attention when researching new interfaces for navigation. Also, as pointed out by McGookin et al. [10], work done on auditory navigation has primarily been geared towards people with visual impairments. There have been, though, a number of efforts developing auditory systems for navigation for sighted. *AudioGPS* by Holland et al. [11] is early work with spatial, non-speech audio to convey information about the direction and distance to a target. *gpsTunes* by Strachan et al. [12] and *Ontrack* [13] by Jones et al. used spatially modified music to convey the same information. *Beowulf* [14] showed that a soundscape together with a low-resolution graphic map is enough to present an entertaining and suitably challenging computer game.

In *Audio Bubbles*, McGookin, Brewster and Priego [10] used audio to inform tourists about nearby points of interest. The users of the system can attend to or ignore the audio information. The aim of the Audio Bubbles is to promote a serendipitous or "stumble upon" type of navigation that is more targeted to exploration and experience than efficiency. The bearing-based navigation used in this study holds the potential to work in a similar way. By leaving the user free to find a suitable route to a selected target herself, she will potentially stumble upon things and locations she did not anticipate being interested in.

HaptiMap [28] has produced a number of results related to the design, implementation and evaluation of maps and location services that are more accessible through the use of several senses such as touch, hearing and vision. See for example [15] and [16]. Suitable angle sizes for pointing gestures were studied in [17]. *SoundCrumbs* [18] uses an interesting navigation method where a trail of virtual "crumbs" is laid out and the application helps users to follow this trail via vibro-tactile cues. The *PointNav* [19] prototype allows a user to both scan for points of interest (POIs) and to get guidance to selected POIs using a combination of pointing gestures, vibro-tactile cues and speech.

The works referred to above have all been successful in using multimodal interfaces to guide users to selected locations. The study described in this article continues this work and adds insights into the attentional and cognitive resources needed when using this approach on navigation.

Surprisingly little of the new opportunities to design multimodal user interfaces have been employed when designing computer games. Some games for smartphones claim to be "pervasive", meaning they depend on and integrate with the environment where they are played [20, 21]. The iPerG project [22] has made extensive research into the area of pervasive gaming.

Djajadiningrat et al. [23] argue that good interaction design should respect all of man's skills: cognitive, perceptual-motor and emotional skills. This leads to interaction design where also what the user perceives with her senses and what she can do with her body become important in the design process. Hekkert [24] divides experience into three levels: aesthetic level, understanding level, and emotional level, and sees aesthetics as "pleasure of the senses". It can generally be argued that aesthetics is a vital part of any user experience and is essential in developing useful, easy to use, and attractive products. The work reported here has strived to embody these ideas in the applications developed.

2.1 Tests performed in this project

Two prototypic smartphone applications were developed, a tourist guide and a gaming application. Both applications had multimodal user interfaces based on GPS positioning and electronic compass data for input from the users, and a balanced mix of spatial and non-spatial audio, text and graphics as output to the users. Both applications were tested in order to assess users' reactions and relationships to using mobile applications with multimodal user interface. Both applications were developed for Android version 2.3 and tested using a mix of qualitative and quantitative methods. During the tests, all subjects used Samsung Galaxy S devices and Koss PortaPro headphones.

3. GUIDE APP FOR VISITOR INFORMATION

The guide app was set up to test how multimodal user interfaces could be used when designing search functions and guide functions for interactive, mobile tourist guides. The search function developed allowed users to search for points of interest (POIs) and the guide function to get guidance to specific locations. The research questions and corresponding hypotheses for design were:

Q1: Can an interface based on a combination of point and sweep gestures, audio feedback and text be used to effectively find information about nearby points of interest?

H1: Users will be able to effectively find information about nearby points of interest using a search method based on a combination of point and sweep gestures, audio feedback and text.

Q2: Can users effectively navigate to specific locations in a city using a guide function that is based on a combination of virtual, spatial sound sources and a graphical arrow to indicate directions to targets and text to indicate distance?

H2: A majority of users will be able to effectively navigate to specified locations in a city using a guide function based on a combination of virtual, spatial sound sources and a graphical arrow to indicate directions to targets and text to indicate distance.

To test the hypotheses a smartphone application for pedestrians called PING! (Point, Interact, Navigate, Go!) was developed. Both the search and the guide function were based on point and sweep gestures for input and sound and simple graphics for output. In the study, the users' task was to first use the search function to find directions to three target locations in the city. Secondly they should use the guide function to also navigate there.

3.1.1 Using the Search Function

When using the search function, the users swept the device horizontally in front of them. For each POI the user pointed at the app played a short sound. In this way the user got an overview of POIs in different directions while concentrating on the environment and without having to look at the screen (figure 1).

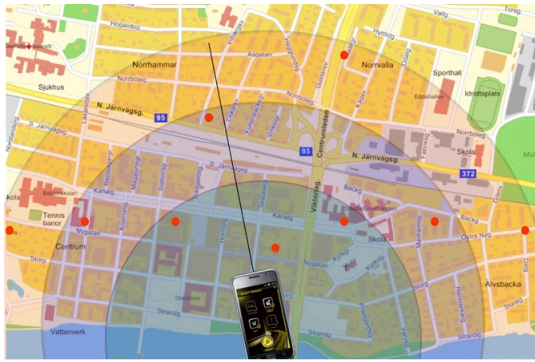


Figure 1. Search function generate a short sound for each POI pointed at

When an interesting direction was identified, the users could get more detailed information about the POIs there by pointing in the direction and tapping an on-screen button. Detailed information was then presented using text and images (figure 2 left and middle).



Figure 2. Left: list of items found in a certain direction. Middle: detail view for one of the items. Right: The guide screen.

3.1.2 Using the Guide Function

The app could also guide users to selected locations through spatial audio, a graphic arrow and text (figure 2 right). The GPS accuracy of smartphones is often not high enough to reliably guide users through turn-by-turn navigation. Instead a method based on the direction to the end target location was developed. The app put a virtual sound source on the target location. Depending on how the user pointed her device in relation to the target, the sound from the virtual sound source was moved between left and right ear. The effect resembled hearing in real-life. The more to the right of the target the user was pointing, the more to the left ear the virtual sound source was moved and vice versa. The users' everyday ability to locate sound sources was in this way used to guide towards the target without the need to interpret more or less abstract information. The direction to the target was also shown with a graphic arrow on the device screen. The distance to the target was encoded into the sound and displayed in digits on the screen. Users were free to choose their personal mix of spatial audio and on-screen information for guidance.

3.2 Guide App Sound Design

The sounds for the guide application were designed to be hearable under varying conditions. They had to be at the same time

pleasant and informative and the user must be able to hear the sounds from the applications over other sounds in the environment. Many sounds in urban environments come from traffic and have their energy spectrums shifted towards lower frequencies [8]. To contrast this, the sounds in the application generally had their energy spectra shifted towards higher frequencies.

Ronkainen et al. [25] have found that minimalist sound design scheme produce better performance. They also found that short sounds in user interfaces made moderate improvement in the performance compared to longer sounds. Therefore the sounds used in the application were kept as short as was deemed possible. Several of the sounds from the application were played many times. It was therefore important to create pleasant sounds with qualities that did not annoy the user. The sounds' attacks were designed to be soft and the timbres not too edgy or cutting. The sounds decay must also be soft. These pleasantness parameters were then balanced against the users' ability to hear the sound.

The user must be able to interpret the sounds in ways consistent with the designer's intention. In this application, many of the sounds had their associated meaning coded into the sounds' left-right position in the stereo image. In other cases they were meant to be very quick and easy to learn from just a short introduction and the context in which they were played.

3.2.1 Sounds For the Search Function

The user searched for POIs by pointing and sweeping the device back and forth. When the system detected a POI in front of the device within a distance of 2 km, the search function played a short sound (< 1 s). The sound came from a plucked guitar string. This sound source was selected for three reasons. First, the guitar is a musical instrument developed for centuries and it can be hypothesised that its sound therefore is acceptable and agreeable to wide user groups. Secondly, the sound was chosen based on its ability to produce short sounds yet with a clearly distinguishable pitch. Lastly the sound from the guitar was chosen since it has an overtone spectrum suitable for low-pass filtering to create a sense of difference in distance.

Three acoustic parameters of the search sound were modulated as a function of the distance to the POI. To make the difference between distances to POIs more clear, the distance was not treated linearly. Instead it was divided in three sections: near, middle and far. Depending on in which section a POI was, the corresponding modulation was applied to the sound. The three acoustic parameters modulated were low-pass filtering, reverberation and pitch (table 1). The more far away a POI was, the more low-pass filtering was applied, the more reverberation was added and the lower its pitch was. This is in accordance with recommendations found in [26]. A side effect of the low-pass filtering is that the sound would appear to be more silent the more filtering that was applied.

Table 1. Modulation parameters for search function sounds

Distance	Low-pass filter	Reverb	Pitch
Near	None	None	F#4 - 369 Hz
Middle	f = 643 Hz Q = 5.6	Middle size room -9 dB below level of direct sound	D4 - 293 Hz
Far	f = 285 Hz Q = 7.0	Large cathedral -5 dB below level of direct sound	A3 - 220 Hz

3.2.2 Sounds For the Guide Function

The guide sound was a virtual sound source placed on the target location. As the user moved and turned, the virtual sound source was panned between left and right ear accordingly in the users' headphones. The result was that the user experienced the sound as coming from the target in very much the same way as if the sound source was real.

In order to be easy to understand and to learn, the guide function used a similar sound design scheme as the search function. Several of the guide sounds' parameters were modulated as a function of the distance from the user to the target and as a function of the bearing from the user to the target.

The guide sounds could be played in two modes, "auto" and "continuous". In auto mode, the guide sound was played with ten seconds of silence between. The sound played in auto mode consisted of two layers, a long sustained bowstring sound and three short vibraphone A4 notes (440 Hz). The sound from the vibraphone was modulated using low-pass filtering and reverb as a function of the distance from the user to the target. As with the search function, the distance between user and target was divided in three zones, "far", "middle" and "near" (figures 3, 4 and 5). Table 2 shows the parameters and settings used for the three zones. In auto mode these sounds were played with 10 second of silence between (figure 6).

Table 2. Modulation parameters for guide function sounds

Distance	Low-pass filter	Reverb	Interval
Near	None	None	285 ms
Middle	$f = 1095 \text{ Hz}$ $Q = 7.0$	Middle size room -12 dB below level of direct sound	780 ms
Far	$f = 312 \text{ Hz}$ $Q = 7.0$	Large cathedral -5 dB below level of direct sound	1200 ms

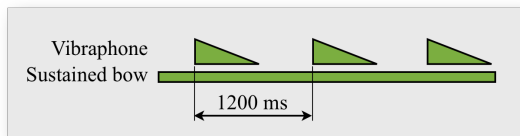


Figure 3. Guide sound for the distance zone "far".

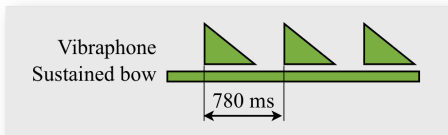


Figure 4. Guide sound for the distance zone "middle".

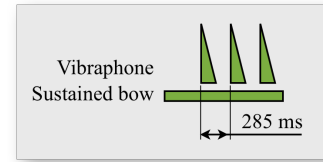


Figure 5. Guide sound for the distance zone "near".

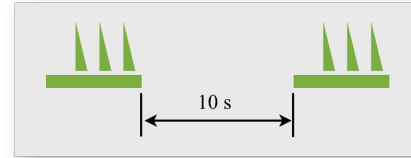


Figure 6. Guide sounds in auto mode, played with 10 seconds intervals.

The guide sound played in continuous mode differed from the sound in auto mode in two respects: there are only the three vibraphone notes played with 2 seconds intervals (figure 7).

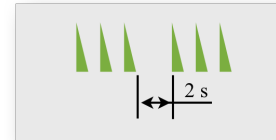


Figure 7. Guide sounds in continuous mode.

3.3 Testing the Guide App

To test the guide application, 24 test users aged 14 to 50 years where enrolled, 15 females and 9 males. For a total time of 90 minutes, the test users were asked to use the app to search for the three specified targets and to use the guide function to walk to these. On each location they should make a note about a specific feature at the location to verify that they had navigated to the correct place. The rest of the time they were asked to use the application to explore the city at their own discretion.

4. RESULTS FROM THE GUIDE APP TESTS

All users of the guide application managed to use the multimodal search function to get information about the three designated targets without problems. All users could also effectively and without problems navigate to the corresponding target locations using the multimodal guide function. It can therefore be argued that both hypotheses H1 and H2 hold true.

4.1 Guide App Tests – Quantitative Results

All test users answered a two-part questionnaire. The first part of the questionnaire was a NASA Task Load Index (TLX) [27]. The second part was six statements with corresponding six grade Lickert scales.

Figure 8 shows the results from the NASA TLX. The graph does not show any high mental or physical loads on the users. Also, the users reported that they thought they succeeded well in navigating the city using the application. They were not overly irritated while doing the task and the application did not hinder their awareness of the surroundings to any great extent while using it.

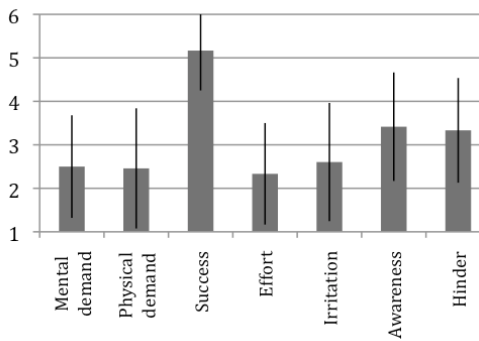


Figure 8. NASA Task Load Index. Average of all users

The second part of the questionnaire contained the following statements.

1. I felt secure and comfortable when using the PING application.
2. The PING application made me feel stressed while moving through the city.
3. While using PING I felt confident that the application would guide me correctly and help me find my way to the targets I selected.
4. When using PING, at several times I felt lost and the application could not help me find what I was looking for.
5. I would like to explore other cities with the help of the PING application.
6. The sound feedback from the PING application was frustrating and confusing.

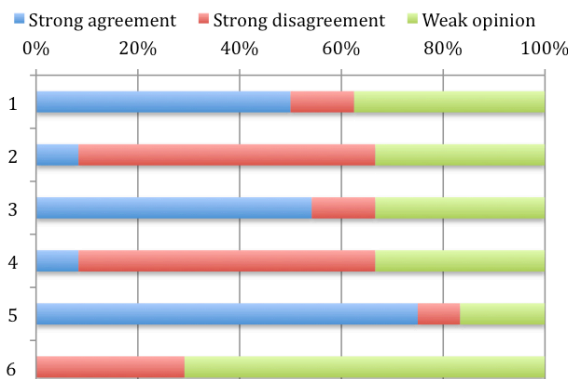


Figure 9. Guide test users opinions' on the questionnaire's complementary statements.

Figure 9 shows the results from the six complementary statements as the percentage of test users in strong agreement with the statements (1 or 2 on the Lickert scales), in strong disagreement with the statements (5 or 6 on the Lickert scales) or showing a weak opinion (3 or 4 on the Lickert scales).

75% of the test users stated that they wanted to explore other cities using the application. No users strongly agreed to the statement that the sound feedback was frustrating and confusing. A majority of the test users reported being confident that the application would help them reach their designated targets.

4.2 Guide App Tests – Qualitative Results

All test users also participated in focus group interviews. Here some users reported using the audio feedback to a large extent,

both when searching for targets and while getting guidance there. These users appreciated being able to use the application more or less “eyes free”. Other users reported relying almost exclusively on the graphical information. Some users asked for greater diversity between the different sounds in order to more easily discriminate between them and their different meanings. One user confused the audio feedback from the application with the ring signal from his mobile phone. At some occasions the sounds from the application was drowned by background noise from traffic or machines. To some users the application did not convey enough information through audio about direction (left/right) or distance to targets. Using speech to give the information “turn left” and “turn right” was suggested as a solution.

Overall, the users reported having relied on the graphical and textual information on the screen more frequently than the information conveyed by the sound feedback. The sound feedback was useful to get information about direction to targets, but in order to get an idea about distance to targets, the users still had to rely primarily on the on-screen text information. The users also reported relying on the on-screen information when the sound from the application was drowned by background noise. Another result is that individual differences in attitudes towards audio, text and graphics were reported to be huge.

No users reported having problems with front/back confusion from the virtual sound source used in the guide function. Pointing the device in slightly different directions, the sound L/R balance changed enough to reassure the users of the correct direction to the target.

5. PERVASIVE GAME APP

The game application built and tested in the same project was named Echo Range. The game used GPS location and pointing direction as input from the users and mainly spatial and non-spatial audio as feedback to the users. The users' locations and movements in the physical world were reflected in the players' avatars moving in the game's virtual world. Information from the virtual game world about directions and distances to opponents was conveyed through audio.

In Echo Range, the users represented warships moving in a virtual game world. The warship's movements in the game world accurately followed the movements of the users in the real world. The users were divided into two opposing teams, submarines and corvettes, with the objective of first seeking out, and then sinking the opposing team's ships. In order to find opponents, users pointed their devices and listened for spatial audio cues. Attacks were made by pointing the device and activating an on-screen “fire-button”. The damage made by an attack was conveyed through audio to both the attacked player and the attacking player. Current health status was shown graphically on the device screen.

In the real world, submarines use stealth and listen passively for corvettes. Corvettes use active sonar and move systematically on the surface searching for submarines. In the game, submarines had a passive, directional listening system based on a 30° cone of sensitivity (figure 10). If a submarine player pointed the device such that a corvette was within the cone, the submarine player heard the virtual engine noise from the corvette. The submarine player then had to decide the likely distance to the corvette. The distance was set with a simple slider on the device screen. A single button was used to fire a virtual torpedo towards the corvette in the pointing direction. When a submarine fired a torpedo, an alarm sounded in the attacked corvette players' headphones. The corvette player then had a few seconds to leave the hit zone before the torpedo arrived and exploded. This is one

example of how a gameplay device can be used to encourage rapid physical movement.

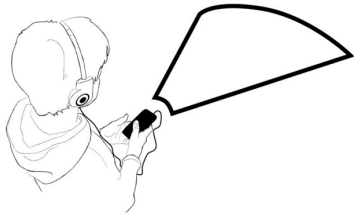


Figure 10. Submarine player listening for engine sounds from corvettes

In contrast, the corvettes used sonar to detect the presence of submarines. Sonars send out short audio pulses and listen for echoes very much like a radar system (figure 11). The corvette players heard a distinct “ping” sound when sending a virtual sonar pulse. For each submarine in the vicinity, the audio pulse was “echoed” back to the sending corvette player. The delay of the returning echo reflected the distance to the submarine, the pitch of the echo reflected the depth of the submarine and the left-right panning of the sound reflected the direction to the submarine. If a submarine was found, the corvette player had to move close enough to the submarine and drop virtual depth charges towards the submarine to destroy it.

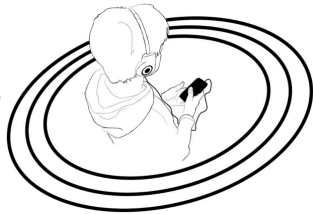


Figure 11. Virtual sonar audio pulse radiating out 360° around corvette player.

The difference in functionalities between the submarine and the corvette were specifically created to encourage two differing trains of tactical thought: On the submarines’ part, it was intended that they should focus on avoidance, passive detection and using the torpedoes as a distance weapon. For the corvette’s, especially as they were armed with a close contact weapon, their tactics involved identifying the probable location of the submarine to target, and then to formulate some strategy for sneaking up on them while, hopefully, in the submarine’s listening blind spot (out of the 30° hydrophone cone).

The research questions and corresponding hypotheses for the game app study were:

Q3: Can a game interface based on a combination of geographic location and pointing gestures as main input from the users and audio and graphics as feedback to the users, give satisfying gaming experiences where the physical world and the virtual game world are perceived as one continuous world?

H3: A game experience based on geographic location and pointing gestures as main input from the users and directional and non-directional audio and graphics as feedback to the users, will

give satisfying gaming experiences where the physical world and the virtual game world are perceived as one continuous world.

Q4: Will users be able to use spatial audio feedback to search for and locate opponents?

H4: A majority of the users will be able to effectively use spatial audio feedback to search for and locate opponents.

5.1 Game App Sound Design

The sounds in Echo Range consisted of a couple of sound families, each having a different and fairly distinct function.

Ambient sounds were designed to give the players an emotional experience of the virtual game environment. In many ways, these sounds replaced the graphic game board or rolling background on which games traditionally rely. The ambient sounds set the scene and continued to support the suspension of disbelief throughout the game experience.

In submarines, the ambient sound consisted of a mixture of the electrics and battery driven engine, the metallic squeaking and creaking of the pressure hull, and the sounds of the ocean depths outside the hull. In the corvettes, the ambient sound consisted of a mix of wind noise, waves against the hull and the diesel engine of the boat itself.

Equipment sounds were designed to give feedback to the players on actions they had taken. For submarine players, the feedback sound was a realistic impression of the releasing of torpedoes. For corvette players the sound was from the releasing of depth charges. These feedback sounds informed the player both as to what action they had just taken and to its eventual, positive or negative, consequences. The sounds were designed to be both realistic and narrative in their nature. The sound sources were a mixture of actual recordings from the vessel-types and layered synthesized sound effects.

Audio was used for the simulated detection systems of both vessels. These systems were used to detect targets or opponents in the game. The submarines’ detection system was passive and was used to listen for engine sounds from corvettes and relied on simulating the real world effects of hydrophonics. Passive listening in real submarines is achieved through an underwater microphone, or hydrophone. The game aims to provide the player with a realistic impression of this by simulating the corvette engine and propeller noise at distance and with the kind of reverberation, diffusion and filtering caused by a body of salt water. The corvette engine sound, heard in the submarines, was compiled of different recordings to provide five distance ranges.

The active sonars in the corvettes were used to detect the presence of submarines. The design of the sonar sound was based on realistic recordings from real sonars layered with soundFXs.

5.2 Testing the Game App

11 users aged 15 to 30 years, three females and eight males, tested the game application. The test users were asked to play the game for about 30 minutes. After the tests, all test users filled out a questionnaire and took part in focus group interviews.

6. RESULTS FROM THE GAME APP TESTS

Although the game application was in an early stage of technical development, the test showed that all players managed to locate opponents by pointing and listening for audio cues. It can therefore be argued that hypothesis H4 holds true. There was also a strong agreement that the players enjoyed playing the game and

that conveying the game world via audio worked according to hypothesis H3.

The test users rated twelve statements on six-level Lickert scales ranging from "Totally agree" (1) to "Totally disagree" (6). The results are shown in figure 12. Strong agreement was defined as answers 1 and 2 on the Lickert scales. Strong disagreement was defined as answers 5 and 6 and weak opinion was defined as answers 3 and 4 on the Lickert scales.

The statements were the following:

1. I enjoyed playing the game
2. I understood how my movements in the real world were reflected in the game.
3. I understood how the game world and the physical world related to each other.
4. I found the game's interface easy to understand and to use.
5. I was able to successfully find an opponent.
6. I was always aware of my real world surrounding.
7. I found myself immersed in the game world.
8. I felt a great sense of achievement when I sunk an opponent.
9. I found the game confusing and too hard to play.
10. I could neither find nor shoot at any opponent.
11. I found the mix of game world and real world confusing.
12. Playing the game disoriented me in the real world.

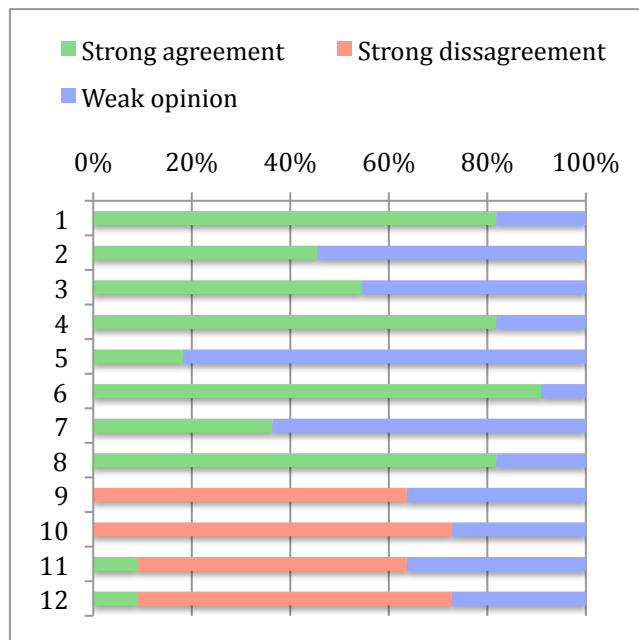


Figure 12. Game test users' opinions on the questionnaire's complementary statements.

Most test users reported enjoying playing the game. No users strongly disagreed to statements 2, 3 and 5 and about half of the users strongly disagreed to statement 11. This indicates that the rendering of the virtual game world mainly through audio did not pose a problem to the players. The results from the focus group interviews support this interpretation.

For games played outdoors and using the surrounding environment as integrated part of the game, the players' safety is an important issue. The reason for rendering the game world in audio was partly to ensure that the players could keep their eyes on the surrounding environment, rather than on a smartphone

screen. The answers to statements 6, 7 and 12 indicate that the users were more or less constantly aware of the physical world surrounding them while playing the game.

7. DISCUSSION

All in all the results show that the multimodal interfaces of the guide and game applications were both at the same time effective and fun to use. The results from the tests show that the multimodal user interfaces with integrated spatial and non-spatial audio cues, contributed significantly to the overall user experience.

Applications featuring multimodal user interfaces hold the potential to unburden users from cognitive loads and interpretative tasks. The results from the tests show that guide applications with such interfaces can potentially help users find points of interest and to guide the way, while at the same time leaving users more free to experience and explore the environment compared to using for example traditional maps. In game applications multimodal interfaces have the potential to support rich pervasive gaming experiences.

Mobile applications often benefit from being possible to use "eyes free", i.e. without forcing the user to constantly keep her eyes on a device screen. Multimodal user interfaces that integrate audio as a bearing part make it possible to design interfaces and user experiences that can be used more eyes free compared to traditional, purely graphical user interfaces.

Sounds from traffic, road worker machinery, people talking loud etc. may from time to time occlude the sounds from multimodal user interfaces and render their sounds impossible or at least hard to hear. In this case it is important to have the same information available for other modalities, primarily vision. This means that the designer of such interfaces should consider presenting the same information simultaneously both in audio and in graphics.

Another reason for considering presenting the same information in multiple modalities is that user preferences differ. Some users need to and/or want to have a graphical user interface, other users only need information and feedback via audio. The same thing probably also affects learning curve effects. Being able to reinforce information on one modality with the same information on another modality may shorten the time needed to learn an interface.

8. ACKNOWLEDGMENTS

The research project was funded by EU Interreg IV A North, the County Administrative Board of Norrbotten, Sweden, the Regional Council of Lapland, Finland and the Interactive Institute, Sweden. Professor Jukka Riekk, Mikko Polojärvi, Timo Saloranta, Mikko Pyykönen, Jari Koloseva and Stefan Lindberg are all acknowledged for their participation in the project.

9. REFERENCES

- [1] A. M. MacEachren. How maps work: representation, visualization, and design. The Guilford Press, 1995.
- [2] Google Street View.
http://maps.google.com/intl/en_us/help/maps/streetview/
- [3] K. Tsukada and M. Yasymua. ActiveBelt: Belt-Type Wearable Tactile Display for Directional Navigation. *Proc. of Ubicomp 04*, Springer, pp. 384-399, 2004.
- [4] M. Frey. CabBoots: shoes with integrated guidance system. *Proc. of the 1st international conference on Tangible and embedded interaction*, ACM pp. 245-246, 2007.

- [5] T. Amemiya, H. Ando and T. Maeda. Lead-me interface for a pulling sensation from hand-held devices. *ACM Trans. Appl. Percept.* 5, 3, Article 15, pp. 1-17, 2008.
- [6] D. Spath, M. Peissner, L. Hagenmeyer and B. Ringbauer. New Approaches to Intuitive Auditory User Interfaces. In *M.J. Smith, G. Salvendy (Eds.): Human Interface, Part I, HCI 2007*, LNCS 4557, pp. 975–984, 2007.
- [7] J. Loomis, J. Marston, R. Golledge and R. Klatzky. Personal Guidance System for People with Visual Impairment: A Comparison of Spatial Displays for Route Guidance. *Journal of Visual Impairment & Blindness* 99, 219–232, 2005.
- [8] R. Kramer, M. Modsching and K. ten Hagen. Development and evaluation of a context-driven, mobile tourist guide. *International Journal of Pervasive Computing and Communications* Vol. 3 No. 4 pp. 378-399, 2007.
- [9] L. Evett, S. Battersby, A. Ridley and D. Brown. An interface to virtual environments for people who are blind using Wii technology - mental models and navigation. *Journal of Assistive Technologies*, Volume 3, Number 2 June 2009, pp. 26-34, 2009.
- [10] D. McGookin, S. Brewster and P. Priego. Audio Bubbles: Employing Non-speech Audio to Support Tourist Wayfinding. In *Proc. HAID 2009*, Springer-Verlag, 41 – 50, 2009.
- [11] S. Holland, D. R. Morse and H. Gedenryd. AudioGPS: spatial audio navigation with a minimal attention interface. *Personal Ubiquitous Computing* 6(4), 253–259, 2002.
- [12] S. Strachan, P. Eslambolchilar and R. Murray-Smith. gpsTunes - controlling navigation via audio feedback. In *Proc. MobileHCI 2005*, vol. 1. ACM, 275–278, 2005.
- [13] M. Jones, S. Jones, G. Bradley, N. Warren, D. Bainbridge and Holmes, G. Ontrack: dynamically adapting music playback to support navigation. In *Personal and Ubiquitous Computing*, 12. Springer-Verlag, 513-525, 2008.
- [14] M. Liljedahl, N. Papworth, and S. Lindberg. Beowulf - An audio mostly game. In *Proc. of International Conference on Advances in Computer Entertainment Technology, ACE '07*, vol. 203, 200-203, 2007.
- [15] D. McGookin, C. Magnusson, M. Anastassova, W. Heuten, A. Rentería, S. Boll (Eds.) *Proceedings from Workshop on Multimodal Location Based Techniques for Extreme Navigation, Pervasive 2010 (Helsinki, Finland)*, 2010.
- [16] M. Anastassova, C. Magnusson, M. Pilot, G. Randall, and G. B. Claassen. Using audio and haptics for delivering spatial information via mobile devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (MobileHCI '10)*. ACM, New York, NY, USA, 525-526, 2010.
- [17] C. Magnusson, K. Rassmus-Gröhn and D. Szymczak. Angle sizes for pointing gestures. In *Proceedings of Workshop on Multimodal Location Based Techniques for Extreme Navigation, Pervasive 2010 (Helsinki, Finland)*, 2010.
- [18] C. Magnusson, B. Breidegard, K. Rassmus-Gröhn. “Soundcrumbs – Hansel and Gretel in the 21st century”, In *Proceedings of the 4th international workshop on Haptic and Audio Interaction Design (HAID '09)*, 2009.
- [19] C. Magnusson, M. Molina, K. Rassmus-Gröhn and D. Szymczak. Pointing for non-visual orientation and navigation. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10)*. ACM, New York, NY, USA, 735-738, 2010.
- [20] Cheok, A.D., Mandryk, R.L., Nilsen, T. Pervasive Games: Bringing Computer Entertainment back to the Real World. *ACM Computers in Entertainment*, Vol 3, No. 3, July 2005, Article 4A
- [21] M. Denward. 2011. Pretend that it is real!: Convergence culture in practice. *Dissertation series in New Media, Public Spheres, and Forms of Expression*, Faculty of Culture and Society, Malmö University. Retrieved 2012-04-27 from <http://dspace.mah.se/bitstream/handle/2043/12240/Denward%20muep.pdf?sequence=2>
- [22] The iPerG Project. <http://www.iperg.org/>
- [23] T. Djajadiningrat, S. Wensveen, J. Frens, and K. Overbeeke. Tangible products: redressing the balance between appearance and action. In *Personal and Ubiquitous Computing 8*. Springer-Verlag, London, 294-309, September 2004.
- [24] P. Hekkert. Design aesthetics: principles of pleasure in design. *Psychology Science*, 48(2), 157-172, 2006.
- [25] S. Ronkainen, J. Häkkinä and L. Pasanen. Effect of Aesthetics on Audio-Enhanced Graphical Buttons. *Proc. ICAD 05*, 2005.
- [26] M. Liljedahl and S. Lindberg. 2011. Sound parameters for expressing geographic distance in a mobile navigation application. In *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound (AM '11)*. ACM, New York, NY, USA, 1-7.
- [27] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Human mental workload*, pages 139–183, 1988.
- [28] The HaptiMap Project. <http://www.haptimap.org/>